ARL-TR-7371 ● AUG 2015

US Army Research Laboratory

An Experimental Exploration of the Impact of Network-Level Packet Loss on Network Intrusion Detection

by Sidney C Smith, Kin W Wong, Robert J Hammell II, and Carlos J Mateo

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TR-7371 ● AUG 2015

US Army Research Laboratory

An Experimental Exploration of the Impact of Network-Level Packet Loss on Network Intrusion Detection

by Sidney C Smith
Computational and Information Sciences Directorate, ARL

Kin W Wong and Carlos J Mateo
ICF International, Fairfax, VA

Robert J Hammell II
Towson University, Towson, MD

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
August 2015	Final	August 2012–April 2015

4. TITLE AND SUBTITLE
An Experimental Exploration of the Impact of Network-Level Packet Loss on Network Intrusion Detection

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Sidney C Smith, Kin W Wong, Robert J Hammell II, and Carlos J Mateo

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
US Army Research Laboratory
ATTN: RDRL-CIN-S
Aberdeen Proving Ground, MD 21005-5067

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-7371

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
In this report we consider the problem of network-level packet loss (NLPL) as it applies to network intrusion detection (NID). We explore 2 research questions: 1) Is there sufficient regularity in NLPL to allow an algorithm to be developed to model it? and 2) Is the impact of network-level packet loss on NID performance sufficiently regular to allow a formula to be developed which will accurately predict the effect? We constructed an experimental environment that mimics the typical placement of an NID sensor. We conducted experiments using MGEN, Pcapreplay, and Snort to explore the impact of NLPL. We discovered that we were unable to produce enough NLPL to characterize its manifestation or analyze its impact on NID.

15. SUBJECT TERMS
computer network security, telecommunication traffic, packet dropper, snort, alert loss rate, network intrustion detection, network traffic composition, packet loss rate

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Sidney C Smith
Unclassified	Unclassified	Unclassified	UU	18	19b. TELEPHONE NUMBER (Include area code) 410-278-6235

Contents

List of Figures

List of Tables

1. Introduction

As we consider the impact of network-level packet loss (NLPL) on network intrusion detection (NID), we observe that NID depends upon the sensor being able to see the traffic between the adversary and the target. General packet loss is very common on the Internet. The Transmission Control Protocol (TCP) is specifically designed to account for general packet loss and uses it as a barometer to gauge the available bandwidth of a connection.[1] In this report, we are not interested in general packet loss because those packets cannot cause a compromise since they will never reach the target. We are interested in what we call detection-relevant packet loss (DRPL). DRPL occurs when packets reach the target but fail to reach the sensor software for analysis. Since the sensors cannot detect what it cannot see, DRPL must have a negative impact on the sensor's ability to detect malicious activity. Based upon the large amount of work that has been done to reduce or eliminate DRPL on NID, we infer that the negative impact of DRPL on NID is well known. This report is part of a larger effort to understand, predict, and model the impact of DRPL on NID. In our previous theoretical work,[2] we divided DRPL into network, host, and sensor levels. In this report, we will focus on DRPL at the network level. NLPL is defined as any packets that proceed to the target but are not sent to the intrusion detection network segment.

The focus of this report is to answer 2 research questions concerning the manifestation of NLPL and the impact of NLPL on NID: 1) Is there sufficient regularity in NLPL to allow an algorithm to be developed to model it? and 2) Is the impact of NLPL on NID performance sufficiently regular to allow a formula to be developed which will accurately predict the effect? We discovered that we were unable to produce sufficient NLPL to explore its manifestation or measure its impact. We concluded that the occurrence of NLPL is so rare as to be negligible.

We provide an overview of the existing research, which is heavily focused upon eliminating packet loss at the network level in Section 2. We discuss our experimental environment, the dataset that we used in these experiments, the experiments to characterize NLPL, and the experiment to measure the impact of NLPL in Section 3. In Section 4, we describe the results of our experiments. Finally, in Section 5 we summarize our results and discuss future work.

2. Background

In previous research,[2] we considered the theoretical impact of packet loss on NID. We divided the potential for packet loss among the network, host, and sensor level. We defined NLPL loss as any packet that reaches the target but fails to reach the network segment where the sensor is located. Additionally, we defined host-level packet loss as any packet that reaches the network interface of the sensor but is not presented to the analysis software. Lastly, we defined sensor-level packet loss as any packet that is presented to the analysis software but is not processed. The various level of packet loss may be seen in Fig. 1. NLPL is represented by bit bucket A. We constructed the packet dropper, which abridges datasets according to several different algorithms we implemented, to emulate our theories about how packets may be lost. We visualized the characterization of this packet loss by graphing the network traffic from 2½ min of traffic from the Cyber Defense Exercise (CDX)[3] dataset, which we describe in detail in Section 3.2, as it was abridged at 25% packet loss by each of the packet-dropping algorithms.[2]

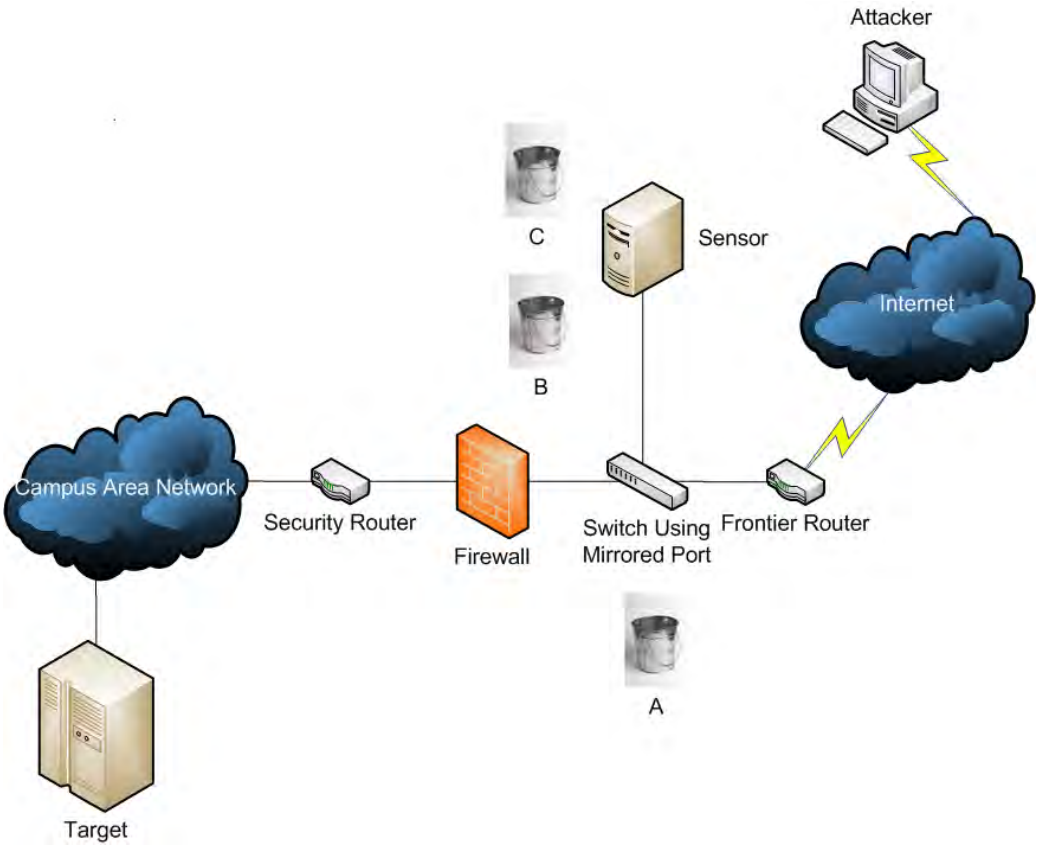

Fig. 1 Breakdown of detection relevant packet loss areas

Although work has been done to reduce or eliminate packet loss, little work has been done to understand and model packet loss and its impact on NID. We were able to glean several insights into packet loss from the work done to reduce or eliminate packet loss. During their work on detecting malicious packet losses, Mizrak et al.[4] encountered the problem of distinguishing malicious packet loss, which is caused by a compromised router from benign packet loss which is simply part and parcel of the way traffic flows through the Internet. They observed that "modern routers routinely drop packets due to bursts in traffic that exceed their buffering capacities, and the widely used TCP is designed to cause such loses as part of its normal congestion control behavior".[4] Although generalized packet loss is not the focus of this research because it is assumed that the target and sensor are seeing the same traffic, many sensors are connected to the network through a mirrored port on a switch. Since mirroring is the lowest priority task that switches perform, there is a possibility that a situation can be created where the malicious traffic would reach the target but fail to reach the sensor.[5] As illustrated in Fig. 1 at bit bucket A, the network path diverges at the mirrored port on the switch. The focus of this research is those packets that reach the target but fail to reach the sensor.

Revisiting Fig. 1 bit bucket A, we see a network switch with 3 network connections: the frontier router, the firewall, and the NID sensor. Given that the routing and firewall functions are significantly more complicated than the switching function, one would expect the switch to easily keep pace with these 2 devices. Unlike network hubs, which are half duplex, network switches are full duplex allowing traffic to flow between the firewall and the frontier router through the switch in both directions at the same time. However, there is still only one traffic path from the switch to the sensor. When packets arriving at the switch from the firewall and frontier router overlap, one must be buffered until the other has completed transmission to the sensor. If there is heavy traffic in both directions, one could imagine this buffer filling and packets failing to be delivered to the sensor.

3. Methodology

Our approach to answering our research questions requires several building blocks. We will need some method to simulate and isolate NLPL. We will use the experimental environment that is described in Section 3.1. Also, we will need NID software to measure the impact of NLPL on NID. We will use Snort[6] for this purpose. Lastly, we will need network capture data in Tcpdump[7] format. We will use portions of the CDX 2009 dataset[3,8] described in greater detail in Section 3.2 for this purpose. We have constructed 2 experiments described in Section 3.3 to attempt to induce NLPL in our experimental environment. Once we were able to

induce 25% packet loss, we planned to capture that data and compare the experimental manifestation of packet loss against our theoretical manifestation of packet loss. Upon completing several trials we planned to process the abridged datasets with Snort using the Community Ruleset. We then planned to graph the packet loss rate against the alert loss rate and compare this graph to similar graphs from our theoretical research.[2]

3.1 Experimental Environment

Figure 2 is a diagram of the network we constructed for conducting our experiments. Table 1 provides the hardware specifications, and Table 2 provides the software specification of this environment. The switch is configured as a layer 3 switch with 2 virtual local area networks (VLANs) and traffic routed between them. The VLANs separate the hosts into an "external" network, VLAN 100, and an "internal" network, VLAN 200. This configuration allows mirroring of all traffic from both VLANs to a collection port which is similar to how the sensors are typically set up.

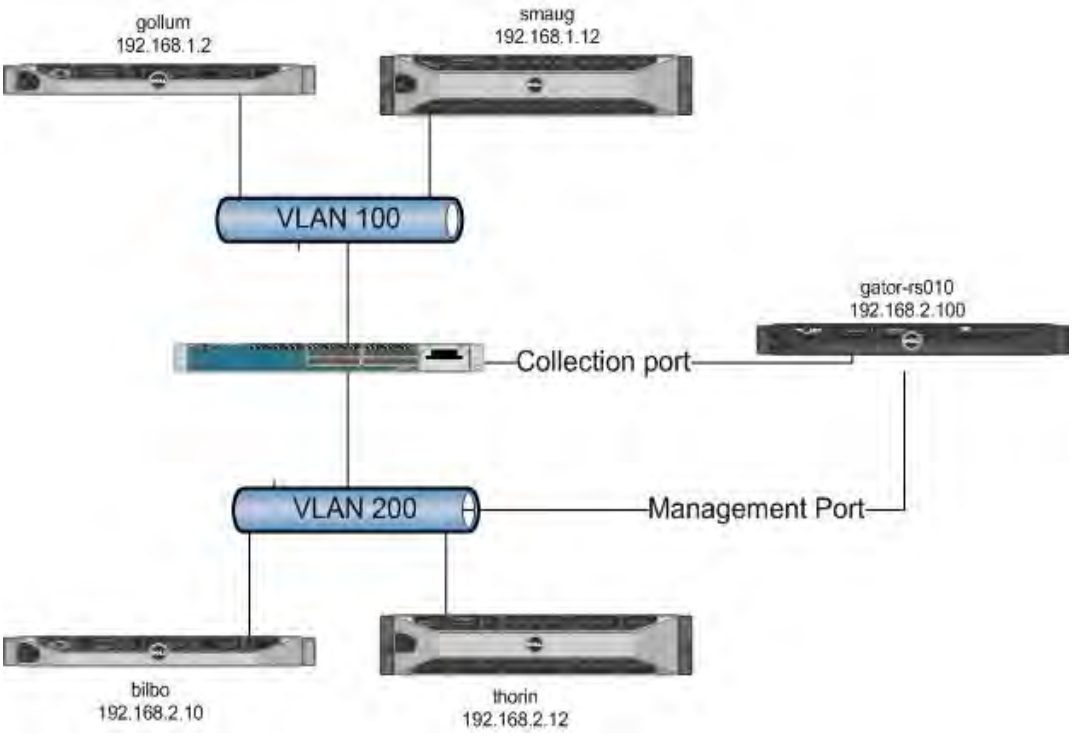

Fig. 2 Experimental environment

Table 1 Hardware specifications

Name	Manufacture	Model	CPU	Memory	Hard Drive	IP Address
Bilbo	Dell	PowerEdge R610	Intel Xeon 16-core X5450 @ 2.53 GHz	12 GB	4× 300 GB 10K SAS	192.168.2.10
Gator-rs010	Dell	PowerEdge R210 II	Intel Xeon 4-core E31220 @ 3.10 GHz	8 GB	. . .	192.168.2.100
Gollum	Dell	PowerEdge R610	Intel Xeon 16-core E5540 @ 2.53 GHz	12 GB	4× 300 GB 10K SAS	192.168.1.2
Smaug	Dell	PowerEdge 2950	Intel Xeon 8-core X5450 @ 3.00 GHz	8 GB	1× 300 GB 10K SAS	192.168.1.12
rsswitch	Cisco	Catalyst 3560-X	. . .	. . .	. . .	. . .
Thorin	Dell	PowerEdge 2950	Intel Xeon 8-core, X5450 @ 3.00 GHz,	8 GB	6× 145 GB 15K SAS	192.168.2.12

Table 2 Software specifications

Name	Source	Version
Snort	www.snort.org	2.9.4
Tcpdump	www.tcpdump.org	4.3.0
Libpcap	www.tcpdump.org	1.3.0
Tcpreplay	tcpreplay.synfin net	3.4.4
MGEN	cs.itd.nrl navy.mil	5.02

3.2 Dataset

Annually the National Security Agency/Central Security Service conducts an exercise pitting teams from the military academies of the United States and Canada against teams of professional network specialists to see who can best defend their network.[8] In their paper “Toward Instrumenting Network Warfare Competitions to Generate Labeled Datasets”, Sangster et al.[3] describe their efforts to collect and label traffic from the 2009 competition. Figure 3 is a diagram of the network used in the competition.

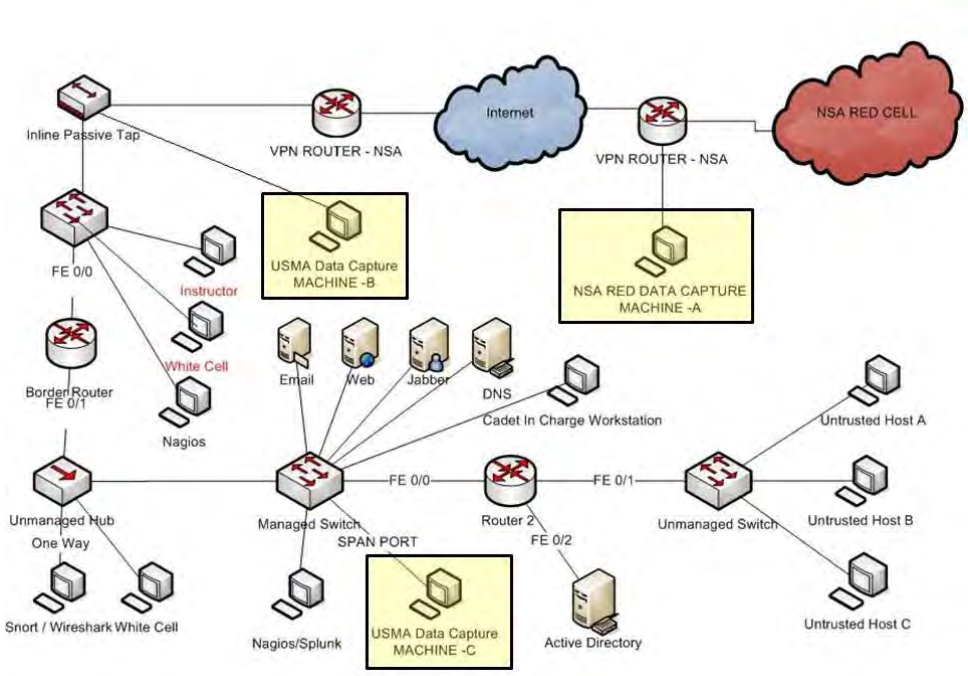

Fig. 3 Instrumented portion of the 2009 Inter-Service Academy Cyber Defense Exercise network[3]

This dataset was chosen because one of the files captured by gator010, 20090421.14.ftrm, contains traffic that is consistent enough for us to be able to see the influence of packet loss on the network traffic. We will be using the same 2 ½ mins of network traffic that was used in our previous research.[2] This will allow us to compare our experimental results with our previous theoretical results.

To use this data in our experimental environment it was necessary to rewrite it using Tcprewrite, which is part of the Tcpreplay[7] package. The traffic had to be split into 2 pieces. One piece contained incoming traffic and the other contained outgoing traffic. The media access control addresses needed to be rewritten to correspond to our experimental environment. This would allow incoming traffic to be switched from VLAN100 to VLAN200 and outgoing traffic to be switched from VLAN200 to VLAN100. This is necessary because only traffic that the switch actually moved from one VLAN to the other would be spanned to the collection port.

3.3 Experiment 1

For our first experiment we configured Gollum to be the MGEN[9] source and Bilbo to be the MGEN sink and sent a steady stream of user datagram protocol packets from Gollum to Bilbo while collecting everything on Gator-rs010. One of the primary purposes of this experiment was to validate that the experimental environment is working correctly.

3.4 Experiment 2

Using Aaron Turner's Tcpreplay,[10] we replayed the same hour of network traffic from the CDX 2009[3] that we used in our theoretical[2] exploration to show the impact of our packet loss algorithms. For this experiment we plugged VLAN100 into one interface on Bilbo and VLAN200 into the other interface on Bilbo and use Tcpreplay to replay the traffic at arbitrary speeds. Table 3 lists the speed multiplier that we used and the packet loss we observed.

Table 3 Network packet loss results for experiment 1

Source Packet Rate (packet/s)	Source Packets Sent	Target Packets Received	Sensor Packets Received	Network Level Packet Loss (%)
40,000	12,000,000	12,000,000	12,000,037	–0.0003
60,000	18,000,000	17,999,840	17,999,872	–0.0002
80,000	24,000,000	24,000,000	23,999,975	0.0001
100,000	30,000,000	30,000,000	30,000,030	–0.0001
120,000	3,600,0000	35,766,571	35,766,602	–0.0001
140000	42,000,000	35,766,571	35,766,601	–0.0001

3.5 Experiment 3

We expect that the switch is most likely to fail when traffic is moving in both directions. Although we replayed traffic in both directions in experiment 2, we were replaying traffic that was captured from a device in a similar configuration to the one we are using. It is possible that this dataset may not have any collisions, because these colliding packets may never have reached the original sensor to be captured. To simulate heavy traffic flowing in both directions, we configured Bilbo and Smaug as MGEN sources and Gollum and Thorin as MGEN sinks and repeated the experiment at several differ bandwidth settings.

4. Results

We were frustrated in our efforts to induce enough packet loss in our experiments to be able to characterize the manifestation of NLPL much less analyze the impact of NLPL on NID.

4.1 Experiment 1

In this configuration we were not able to send enough packets over the switch to cause the mirror to fail. Looking at Table 4, we find that even when the switch was unable to transmit all of the packets that the source was sending, it was still able to span all of the packets that sink received to the sensor.

Table 4 Network packet loss results for experiment 2

Run	Multiplier	Time (sec)	TimeRatio	PktsReceived	PktLoss
1	200	17.73	0.985	1,340,209	0
2	250	14.22	0.988	1,340,212	0
3	300	11.93	0.994	1,340,212	0
4	600	6.43	1.072	1,340,212	0
5	1,000	4.53	1.258	1,340,212	0
6	1,200	4.19	1.397	1,340,245	0
7	1,400	3.94	1.532	1,340,212	0

4.2 Experiment 2

In Table 4 we see the results of replaying the hour of CDX data using Tcpreplay at various speeds. We were able to replay the hour of the CDX 2009 data at speeds—over 1,000 times the original speed—and were not able to produce packet loss at the switch.

4.3 Experiment 3

We ran this configuration at bursts of 30 MB/s, and we were unable to cause the switch to fail. We ran this configuration at bursts of 70 MB/s and saw 5% packet loss. This means that our gigabit switch failed to mirror 5% of the traffic when we pushed 1.12 Gb over the network.

5. Conclusions

We conclude that although it is theoretically possible for a switch to fail to span a packet, at least for the equipment used in a configuration typical for network intrusion detection, NLPL is not a significant problem. As we consider the issues, this really is not that surprising. The spanning switch is part of the larger network security stack (NSS). The front face of the NSS is the frontier router. Routing is significantly more resource intensive than switching. The rear face of the NSS is a firewall. Firewalling is significantly more resource intensive than switching. Lastly wide area network (WAN) bandwidth is typically significantly more expensive than local area network bandwidth. This means that bandwidth capacity available from the frontier router to the Internet is less than the bandwidth capacity between the frontier router and the firewall. The WAN bandwidth capacity serves as a cap for the total amount of traffic that will be processed through the NSS. In this configuration an appropriately sized switch should have no trouble spanning the traffic in the NSS which should be significantly less than the traffic level for which the switch is rated. This finding does not hold in other applications. For example,

our research does not address the use case where the switch is used in the interior of the network to connect several hosts at the same bandwidth level as the spanning network.

The ultimate goal is to discover a general function $y = f(x)$ where y is alert loss rate and x is packet loss rate. This function will allow us to accurately predict the impact of packet loss on NID. To achieve this goal, future research must be conducted on the studies of the experimental impact of host and sensor-level packet loss on NID. Also, a study of the combined effect of host and sensor-level packet loss should be conducted upon the foundation that has been laid by studying each of the layers individually. The packet dropper algorithm needs to be refined and validated based upon the findings of the combined study. Replaying a dataset at several multiples of its original speed is a time consuming process even when it may be completely automated. The CDX dataset that we used was originally 105 h long. Replaying the dataset in an exponential progression (i.e., 2^n) would require 210 h. Often this technique does not produce enough data points and the process would take even longer. Some datasets studied in our previous research are 7 weeks long and would be completely impractical to study by replaying them at several multiples of the original speed. Having a validated packet dropper, which can process the dataset in minutes rather than weeks, would greatly benefit the research. With a validated packet dropper it may be possible to analyze several datasets and generate and validate a general function.

6. References

1. Stevens WR. TCP slow start, congestion avoidance, fast retransmit, and fast recovery algorithms. Freemont (CA): Internet Engineering Task Force; 1997 Jan. RFC No.: 2001.

2. Smith SC, Hammell RJ, Parker TW, Marvel LM. A theoretical exploration of the impact of packet loss on network intrusion detection. In: Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2014 15th IEEE/ACIS International Conference; 2014 Jun 30–Jul 2; Las Vegas, NV. Piscataway (NJ): IEEE; 2014. p. 1–6.

3. Sangster B, O'Connor T, Cook T, Fanelli R, Dean E, Adams WJ, Morrell C, Conti G. Toward instrumenting network warfare competitions to generate labeled datasets. In: Proc. of the 2nd Workshop on Cyber Security Experimentation and Test (CSET09); 2009 Aug 10–14; Montreal, Canada.

4. Mizrak AT, Savage S, Marzullo K. Detecting malicious packet losses. Parallel and distributed systems, IEEE Transactions. 2009; p. 191–206.

5. O'Neill T. SPAN Port or TAP? CSO Beware. LoveMyTool. 2007 Aug 23 [accessed 2012 Feb 21]. http://www.lovemytool.com/blog/2007/08/span-ports-or-t.html.

6. Roesch M. Snort: lightweight intrusion detection for networks. Proceedings of LISA '99. 13th USENIX Systems Administration Conference; 1999 Nov. 7–12; Seattle, WA. Berkley (CA); USENIX Association; 1999; p. 229–238.

7. Turner A, Bing M. Tcpreplay: Pcap editing and replay tools for *nix. synfin 2005 [accessed 2013 Apr 2] http://tcpreplay.synfin.net.

8. West Point takes the NSA cyber defense trophy for the third straight year. Natioinal Security Agency Central Security Service. 2009 Apr 28 [accessed 2013 Mar 22]. http://www.nsa.gov/public_info/press_room/2009/cyber_defense_trophy.shtml.

9. PROTEAN. MGEN User's and Reference Guide. Ver. 5.0. Washington (DC): Naval Research Laboratory (US). [accessed 2015 Apr 20]. http://downloads.pf.itd.nrl.navy.mil/docs/mgen/mgen.html.

10. Tcpreplay. Syn Fin dot Net [accessed 2013 Apr 2]. http://tcpreplay.synfin.net.

List of Symbols, Abbreviations, and Acronyms

CDX	Cyber Defense Exercise
DRPL	detection-relevant packet loss
NID	network intrusion detection
NLPL	network-level packet loss
NSS	network security stack
TCP	Transmission Control Protocol
WAN	Wide Area Network

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
2 (PDF)	DIRECTOR US ARMY RESEARCH LAB RDRL CIO LL IMAL HRA MAIL & RECORDS MGMT
1 (PDF)	GOVT PRINTG OFC A MALHOTRA
1 (PDF)	DIR USARL RDRL CIN S S SMITH